Connected Tropical Subgraphs in Vertex-Colored Graphs

J.-A. Angles d'Auriac*, N. Cohen*, A. El Maftouhi*,
A. Harutyunyan*, S. Legay*, Y. Manoussakis*
*University Paris South, L.R.I., Bt. 650, 91405 Orsay Cedex, France.

July 3, 2014

In this work, we deal with tropical substructures in vertex-colored graphs, first introduced in [1]. Vertex-colored graphs are useful in various situations. For instance, the Web graph may be considered as a vertex-colored graph where the color of a vertex represents the content of the corresponding page (red for mathematics, yellow for physics, etc.) [2]. Applications can also be found in bioinformatics (Multiple Sequence Alignment Pipeline or for multiple protein-protein Interaction networks) and in data bases (e.g. online social networks or matching products in online stores)[4]. Given a vertex-colored graph, a *tropical subgraph* (induced or not) is defined to be a subgraph where each color of the initial graph appears at least once. Potentially, any kind of usual subgraphs (paths, cycles, independent sets, dominating sets, vertex covers, etc.) can be studied in their tropical version. Here, we study minimum connected tropical subgraphs in vertex-colored graphs.

Throughout this abstract, we let $G = (V, E)$ denote a simple undirected graph, and G^c a (not necessarily properly) vertex-colored simple undirected graph with c different colors. A connected subgraph H of G^c is said to be *tropical* if each color of G^c is present at least once in H. The *connected tropical subgraph number* $\mathfrak{tc}(G^c)$ is the order of a smallest connected tropical subgraph of G^c. A *connected rainbow subgraph* of G^c is a connected subgraph in which each color is present at most once. A *connected colorful subgraph* of G^c is a connected rainbow subgraph which is tropical. We let $\delta(G^c)$ denote the minimum degree of G^c. In this work, we study the following problem, from both an algorithmic and mathematical point of view:

Minimum Connected Tropical Subgraph Problem (MCTS)

Input: Vertex-colored graph G^c
Question: Determine the value $\mathfrak{tc}$.

In a first series of results below, we deal with NP-hardness, as well as sufficient conditions for MCTS.

Theorem 1. *MCTS is NP-Hard on trees, interval graphs and split graphs.*

We obtain these results by reduction from Dominating Set, Vertex Cover and Vertex Cover, respectively.

Theorem 2. *Given a vertex-colored graph G^c, MCTS can be solved on G^c in $O(n^2 \times m \times 8^c)$ time.*

We now provide sufficient conditions for the existence of a connected tropical subgraph of a given order.

Theorem 3. *Let G^c be a connected vertex-colored graph with n vertices and m edges. For each $k \in \mathbb{N}$, if $m \geq \binom{n-k-2}{2} + n - c + 2$, then $\mathfrak{tc}(G^c) \leq c + k$.*

Theorem 4. *Let G^c be a vertex-colored graph of minimum degree δ. If $\delta \geq \frac{n}{2}$ and $c \geq \frac{n}{2}$, then G has a connected colorful subgraph, and we can find it in polynomial time.*

The bounds of both theorems are tight as we can provide extremal graphs. In addition, the next theorem shows that Theorem 4 cannot be significantly improved.

Theorem 5. *Let $r\delta(G^c)$ be the rainbow degree, i.e., the smallest number of colors a vertex of G^c can have in its neighborhood.*

1. *Let $\epsilon, \epsilon' \in [0, 1)$. There exists a vertex-colored graph G^c such that $\delta(G^c) \geq \epsilon n$, $r\delta(G^c) \geq \epsilon' c$ and G^c has no connected colorful subgraph.*

2. *Let p be an integer. There exists a vertex-colored graph G^c such that $\delta(G^c) \geq n - c + p$ and G^c has no connected colorful subgraph.*

3. *Let p be an integer, and $\epsilon \in [0, 1)$. There exists a vertex-colored graph G such that $\delta \geq \epsilon n$, G is p-connected and has no connected colorful subgraph.*

In the rest of our work, we study MCTS in the case of vertex-colored random graphs. We recall that the random graph $G(n, p)$ is the graph on n vertices where each of the possible edges appears with probability p, independently. Given a positive integer c, let $G(n, p, c)$ be the graph obtained from $G(n, p)$ by coloring each vertex with one of the colors $1, 2, \ldots, c$ uniformly and independently at random. In what follows, we will say that $G(n, p, c)$ has a property $\mathcal{Q}$ *asymptotically almost surely* (abbreviated a.a.s.) if the probability it satisfies $\mathcal{Q}$ tends to 1 as n tends to infinity. The threshold function for the property of containing a copy of a fixed graph G is $n^{-1/m(G)}$ where $m(G)$ is the ratio of the number of edges to the number of vertices in the densest subgraph of G, that is,

$$m(G) = \max \left\{ \frac{e_H}{v_H} \ : \ H \subseteq G \ , \ v_H > 0 \right\},$$

where v_H and e_H stand for the number of vertices and edges of H, respectively. The following theorem shows that this threshold also holds for the property of the existence of a tropical copy of a given graph in $G(n, p, c)$. The proof of this result is fully based on that of the theorem of Bollobás.

Theorem 6. *Let G be a fixed graph with at least one edge, $e_G > 0$. Let $c = v_G = |V(G)|$. Then*

$$\lim_{n\to\infty} \mathbb{P}\left[G(n,p,c) \supset \text{tropical copy of } G\right] = \begin{cases} 0 & \text{if } p \ll n^{-1/m(G)} \\ 1 & \text{if } p \gg n^{-1/m(G)}. \end{cases}$$

In the next theorem we investigate the case in which $pn^{1/m(G)} \to \theta$ as $n \to \infty$, where θ is a positive constant. We are specially interested in a family of graphs called strictly balanced graphs defined as follows. A graph G is *balanced* if $m(G) = e_G/v_G$, that is, if $e_H/v_H \leq e_G/v_G$ for every $H \subset G$. G is *strictly balanced* if $e_H/v_H < e_G/v_G$ whenever $H \subsetneq G$, that is to say that every proper subgraph of G is strictly less dense than the graph itself. Trees, cycles and complete graphs are strictly balanced.

Theorem 7. *Let G be a fixed strictly balanced graph with v vertices and e edges. Denote by $a = |Aut(G)|$ the number of elements of the automorphism group of G. Let θ be a positive constant and set $p = \theta/n^{v/e}$. Let $X_n(G)$ denote the number of tropical copies of G in $G(n,p,c)$ with $c = v$. Then*

$$X_n(G) \xrightarrow{d} \mathcal{P}(\lambda) \quad \textit{with} \quad \lambda = \frac{c!\theta^e}{ac^c},$$

where $\mathcal{P}(\lambda)$ is the Poisson distribution with mean λ.

One of the most interesting results in the study of random graphs was discovered by Matula [3] who proved that the clique number $\mathrm{cl}(G(n,p))$ of $G(n,p)$ is asymptotically almost surely concentrated on two consecutive values. Let $0 < p < 1$ be fixed and set $b = 1/p$. Let the function $d(n)$ be defined by

$$d(n) = 2\log_b n - 2\log_b \log_b n + 1 + 2\log_b (e/2)\,.$$

Then, for any $\epsilon > 0$, the clique number of $G(n,p)$ satisfies

$$\Pr\Big[\lfloor d(n) - \epsilon \rfloor \leq \mathrm{cl}\big(G(n,p)\big) \leq \lfloor d(n) + \epsilon \rfloor\Big] \to 1 \quad \textit{as} \quad n \to \infty.$$

This leads us to the natural question of what is the maximum number of colors $c = c(n)$ which a.a.s. guarantees the existence of a tropical clique of order r in $G(n,p,r)$, for every $r \leq c(n)$. The answer to this question is given by the following theorem.

Theorem 8. *Let $0 < p < 1$ be fixed. Let $c = c(n)$ be the function defined by*

$$c(n) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2 + 1,$$

where $b = 1/p$.

(i) If $r > c(n)$, then a.a.s. there is no complete tropical subgraph of order r in $G(n,p,r)$

(ii) If $r < c(n)$, then a.a.s. $G(n, p, r)$ contains a complete tropical subgraph of order r.

The next theorem finds the distribution of T_k for certain values of k.

Theorem 9. *Let $p = \theta/n$, where $0 < \theta < 1$ is fixed. Let*

$$k = \frac{1}{\theta - \log\theta}\Big[\log n - 2\log\log n + l\Big],$$

where l is a fixed real number. Denote by T_k the number of components of $G(n, p, k)$ that are tropical trees of order k. Then T_k has asymptotically Poisson distribution $\mathcal{P}(\lambda)$ with mean

$$\lambda = \frac{(\theta - \log\theta)^2 e^l}{\theta}.$$

References

[1] J.-A. Anglès d'Auriac, H. El Mafthoui, M. Karpinski, Y. Manoussakis, L. Montero, N. Narayanan, L. Rosaz, and J. Thapper. Tropical dominating sets in vertex-colored graphs. Submitted.

[2] S. Bruckner, F. Hffner, C. Komusiewicz, and R. Niedermeier. Evaluation of ilp-based approaches for partitioning into colorful components. In *Software Engineering and Applications*, pages 176–187, 2013.

[3] D. W. Matula. The largest clique size in a random graph. *Technical Report, Dept. of Computer Science, Southern Methodist University Dallas*, 1976.

[4] C. Zheng, K. Swenson, E. Lyons, and D. Sankoff. Omg! orthologs in multiple genomes - competing graph-theoretical formulations. In *Workshop on Algorithms in Bioinformatics*, pages 364–375, 2011.